

学校编码: 10384

分类号\_\_\_\_\_密级\_\_\_\_\_

学 号: 10220070153976

UDC\_\_\_\_\_

厦 门 大 学

博 士 学 位 论 文

# 面向词义自动标注的多义动词选择限制研究

Research on Selectional Restriction of Polysemous Verb  
for Word Sense Tagging

蒋媛

指导教师姓名: 苏新春教授

专 业 名 称: 汉语言文字学

论文提交日期: 2015 年 4 月

论文答辩日期: 2015 年 5 月

学位授予日期:

答辩委员会主席: \_\_\_\_\_

评 阅 人: \_\_\_\_\_

2015 年 6 月



---

## 厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下,独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果,均在文中以适当方式明确标明,并符合法律规范和《厦门大学研究生学术活动规范(试行)》。

另外,该学位论文为( )课题(组)的研究成果,获得( )课题(组)经费或实验室的资助,在( )实验室完成。(请在以上括号内填写课题或课题组负责人或实验室名称,未有此项声明内容的,可以不作特别声明。)

声明人(签名):

年 月 日



## 厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

（        ） 1. 经厦门大学保密委员会审查核定的保密学位论文，于        年        月        日解密，解密后适用上述授权。

（        ） 2. 不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年        月        日



## 摘要

本研究的构想在课题“多义词词义搭配知识库与词义的自动标注（SCT）”的理论探讨和实际操作中形成。研究选取了《现代汉语词典》第3版（1996）中903个双音节、义项数为2-5个的多义动词（共计2002个义项），对其区别性搭配进行了提取。结合选择限制在词义自动标注领域的应用，分析了多义动词选择限制的表现、特点、机制，并对词义自动标注中遇到的问题进行了探讨，针对词义自动标注的实际对机用义项库进行了调整。

无论词义自动标注采用什么样的知识源、何种方法，最根本的依据是多义词上下文中的区别性形式特征（以语义搭配为主）。区别性形式特征是语言单位之间选择限制的结果。选择限制基于义项间的“对立”与“差别”，是一种普遍存在于语言系统中的机制。选择限制在词义自动标注领域有重要应用。知识源是词义自动标注的关键，解决知识源问题需要回归到语言本身，对选择限制的规律、机制进行考察。义项形式论是本研究的基础理论，区别性形式特征是本研究的核心概念。研究分为六个部分来阐释上述内容：

第一章介绍了选题缘由、选题意义、研究思路、研究材料、研究方法，并对术语进行了说明。

第二章回顾了词义自动标注领域与多义动词选择限制领域的相关研究，指出在选择限制知识的获取、理论方法的进步、机用词典的建立、选择限制分析模式的形成、选择限制规律与机制的发掘等方面还需进一步研究。

第三章讨论了多义动词选择限制知识的获取与形式化。形式化是词义自动标注的关键环节，多个语言资源通过这一环节构成词义标注的有机系统。义项存在的价值在于义项之间的对立与差别，这种对立与差别使义项具有了形式特征，这是词义自动标注的前提；区别性的形式特征可以被“形式化”为“规则”，这是词义能够被自动标注的依据。区别性形式特征主要体现为搭配，语义搭配是“形式化”的主要对象。“形式化”是对区别性形式特征性质和功能的抽象，实现了由自然语言向人工语言的转换，得到了描述语言关系、规律、机制，集中体现词语间选择限制的“规则”。

第四章分析了多义动词选择限制在语法、语义层面的表现，总结了多义动词选择限制的特点，探索了选择限制的规律与机制。指出词义关系的对立具有先决性，在词

义自动标注领域,词汇意义层面的对立比色彩意义层面的对立易于标注,表达不同事件的对立比表达同一事件的对立易于标注。另外,多义动词的选择限制具有不平衡性、系统性的特点,选择限制的力度有强弱之分。选择限制力的强弱与动词动作性的强弱、词义的限定元素有关。动作性的强弱可以通过语义特征 [±自主]、[±及物]、[±有生]、[±开始]、[±完成]、[±持续]、[±动作]、[±行为]、[±变化]、[±状态]、[±关系]来衡量。认知理论也为选择限制力的强弱提供了解释。

第五章结合多义动词规则的验证,探索了多义动词机用义项库的建设问题。验证是对规则有效性、义项设置合理性的检视,是词义自动标注的重要环节。验证结果显示已有规则具备有效性,同时 SCT 机用义项库需要进一步调整。义项的设置应具备独立性、排他性、概括性,要在语料统计的基础上,以语境中的区别性形式特征为线索,以核心语义特征为依据进行处理。义项库的调整应形成科学、统一的标准。义项的粒度可以因词而异。

第六章对研究的内容进行了梳理与概括,并指出研究的不足与进一步研究努力的方向。

**关键词:** 多义动词; 选择限制; 区别性形式特征



## Abstract

The conception of this research crystallized in the theoretical discussion and practical operation on “*Sense Collocation Base and Word Sense Tagging (SCT)*”. 903 disyllabic polysemous verbs (2-5 senses) and 2002 senses are involved in the research, and their distinctive collocations have been abstracted. The author analyzed the manifestations, characteristics, mechanism and issues of the polysemous verb selectional restriction, which combined with the application of selectional restriction in word sense tagging. And the author explored the establishment of word sense base for computer, which is based on the actual conditions of word sense tagging.

Word sense tagging is based on the distinctive formal features (distinctive collocations are in the majority) from context, whichever the source of knowledge or the method has been chosen. The distinctive formal features stem from selectional restriction. Selectional restriction is a kind of general mechanism in linguistic system, which comes from the opposition and difference among senses. The selectional restriction plays an important role in word sense tagging, and the source of knowledge is the linchpin of this field. To solve the problem requires returning to linguistics, to probe into the regularities and mechanism of selectional restriction. Formal feature of sense theory is the cornerstone of this research, and “distinctive formal feature” is the core concept. The article is divided into six chapters.

Chapter one mainly introduces the basis of topic selection, significances, research methods, research materials, and related-terminology of this article.

Chapter two reviews the research status of word sense tagging and polysemous verb selectional restriction, points out the need for further research about the abstracting of selectional restriction knowledge, the progress of methods, the establishing of word senses base and the analytical model of selectional restriction, in addition, the analysis about regularities and mechanism of selectional restriction.

Chapter three interprets the abstracting and formalizing of polysemous verb selectional restriction knowledge. Formalizing plays a key role in word sense tagging, and links language resources bases. The value of sense stems from the opposition and

difference with each other, and the distinctive formal features base on the opposition and difference, which is the premise of word sense tagging. The distinctive formal features can be “formalized” to “rules”, which is the basis of word sense tagging. The distinctive formal features mostly manifest themselves as collocations. The “formalization” targets at semantic collocations mainly, which is an abstraction from the characters and functions of distinctive collocations, and transforms natural language to artificial language. The rules describe the relationships, principles and mechanism, and concentrate on the selectional restriction among words.

Chapter four analyses the manifestations of polysemous verb selectional restriction on syntax and semantics, sums up the characteristics, and explores the regularities and mechanism of polysemous verb selectional restriction. On the field of word sense tagging, the opposition among senses is primacy. Tagging lexical meaning opposition is more easily than color meaning opposition, and tagging opposition which represent different events is more easily than which represent same event. The characteristics of polysemous verb selectional restriction are systematicness and unbalance, and the performance of selectional restriction is different. The power of selectional restriction is affected by the quality of verb and the semantic component. The quality of verb can be measured by a series of semantic features, which are [ $\pm$ independent], [ $\pm$ transitive], [ $\pm$ animate], [ $\pm$ beginning], [ $\pm$ completed], [ $\pm$ sustained], [ $\pm$ motional], [ $\pm$ active], [ $\pm$ variational], [ $\pm$ stative] and [ $\pm$ relative]. Cognitive Linguistics provides some interpretations for the performance of selectional restriction.

Chapter five explores the establishing of polysemous verbs senses base, which is united with the test and verifying of polysemous verbs rules. The test and verifying aims at the effectiveness of rules and the appropriateness of senses, and it is an important part for word sense tagging. The result reflects that rules are effective, meanwhile, the word senses base needs to be adjust. All senses from the base should possess independence, exclusiveness and generality. In order to achieve the objective, the setting of senses should be based on the authentic language material, the distinctive formal features, the core of semantic characteristics, and scientific standards. The granularity of sense differs from each other.

Chapter six summarizes the content of the research, and points out the deficiencies and future research directions.

**Keywords:** polysemous verb; selectional restriction; distinctive formal features

厦门大学博硕士学位论文摘要库



## 目录

<b>第一章 绪论</b>	<b>1</b>
<b>第一节 研究缘起</b>	<b>1</b>
一、选题的实践基础	1
二、选题的现实与理论依据	3
三、选题意义	6
<b>第二节 研究说明</b>	<b>7</b>
一、研究定位和任务	7
二、研究对象	8
三、研究资源	10
四、研究方法	11
五、研究设计	12
六、术语说明	12
<b>本章小结</b>	<b>15</b>
<b>第二章 相关研究综述</b>	<b>16</b>
<b>第一节 研究的理论基础</b>	<b>16</b>
一、语言系统论	16
二、义项形式论	17
<b>第二节 词义自动标注研究综述</b>	<b>20</b>
一、概况	20
二、词义自动标注的方法	21
三、国内的词义自动标注研究	25
四、对词义自动标注研究的小结与思考	26
<b>第三节 多义动词的选择限制研究综述</b>	<b>28</b>
一、“选择限制”问题的提出	28
二、选择限制的理论研究	29
三、多义动词的选择限制研究	33
<b>本章小结</b>	<b>39</b>

**第三章 多义动词选择限制知识的获取与形式化 .....41****第一节 选择限制知识的描写：SCT 多义动词搭配知识库的框架设计.....41**

- 一、SCT 多义动词搭配知识库框架设计的理论依据 .....41
- 二、SCT 多义动词搭配知识库框架的设计原则 .....45
- 三、SCT 多义动词搭配知识库框架的结构 .....46
- 四、SCT 多义动词搭配知识库框架的字段说明 .....47
- 五、对 SCT 多义动词搭配知识库框架设计的反思 .....55

**第二节 选择限制知识获取的基础：SCT 语言资源的构建.....56**

- 一、SCT 现代汉语综合语料库的构建 .....56
- 二、SCT 现代汉语义项库的构建 .....57
- 三、SCT 现代汉语语义分类库的构建 .....58
- 四、对 SCT 现代汉语语法库的利用 .....60

**第三节 选择限制知识的形式化：多义动词搭配规则库的形成 .....62**

- 一、多义动词选择限制知识形式化的目标.....62
- 二、多义动词选择限制知识形式化的实例.....63
- 三、多义动词选择限制知识形式化的问题.....66

**本章小结.....69****第四章 多义动词选择限制的表现、特点与机制 .....71****第一节 多义动词在语法层面的选择限制表现 .....71**

- 一、动词语法功能的相关研究.....71
- 二、多义动词对特定词的选择限制.....76
- 三、多义动词对宾语的选择限制.....78
- 四、非句子中心的多义动词对其句法搭配的选择限制.....81
- 五、多义动词句法层面选择限制分析对词义自动标注的意义.....83

**第二节 多义动词在语义层面的选择限制表现 .....86**

- 一、语义层面的多义词研究.....86
- 二、多义动词的义项对立关系类型.....90
- 三、多义动词语法层面选择限制差异的语义分析.....93
- 四、多义动词语义选择限制分析对词义自动标注的意义.....95

<b>第三节 多义动词选择限制的特点 .....</b>	<b>96</b>
一、多义动词选择限制的不平衡性 .....	96
二、多义动词选择限制的系统性 .....	102
三、词义自动标注对多义动词选择限制特点的利用 .....	102
<b>第四节 多义动词选择限制的机制 .....</b>	<b>103</b>
一、从词义的角度考量选择限制的机制 .....	103
二、从认知的角度考量多义动词选择限制的机制 .....	104
<b>本章小结 .....</b>	<b>105</b>
<b>第五章 规则的验证与 SCT 多义动词义项库的调整 .....</b>	<b>107</b>
<b>第一节 SCT 多义动词规则的验证 .....</b>	<b>107</b>
一、多义动词验证库的设计 .....	107
二、多义动词的验证结果 .....	108
三、多义动词规则验证结果的分析 .....	109
<b>第二节 SCT 多义动词义项库的调整 .....</b>	<b>112</b>
一、SCT 机用义项库与传统词典的共性和差异 .....	113
二、SCT 多义动词义项库的调整 .....	120
<b>本章小结 .....</b>	<b>126</b>
<b>第六章 结语 .....</b>	<b>128</b>
一、结论与本文创新之处 .....	128
二、不足之处与余论 .....	130
<b>参考文献 .....</b>	<b>131</b>
<b>附录： 903 个动词词目 .....</b>	<b>141</b>
<b>附录： 词性对照表 .....</b>	<b>144</b>
<b>附录： 后记 .....</b>	<b>145</b>
<b>附录： 读博期间科研成果 .....</b>	<b>147</b>





Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to [etd@xmu.edu.cn](mailto:etd@xmu.edu.cn) for delivery details.

厦门大学博硕士论文摘要库